Executive Summary

Susan M. Ascher, MD, Faina Shtern, MD

The emergence of national and state health reforms, health care cost containment, and the continued shift to managed care and evidence-based medicine, concurrent with new medical advances, underscore the need to incorporate technology assessment research into the clinical decision-making process. No serious policy, strategy, or programmatic initiative regarding new technologies can be accurately formulated without reliable clinical and economic data.

The currently dominant paradigm for assessing the effects of medical care focuses on the evaluation of therapeutic technologies. The evaluation of diagnostic imaging in this process is less well developed. Traditionally, most evaluation of radiologic examinations has been empirical, consisting of retrospective or prospective analyses of their sensitivity and specificity. In most instances, there has been no linkage of the sensitivity and specificity data with clinical decisions, patient outcomes, and cost-effectiveness. Research synthesis studies have not gained widespread acceptance or implementation. In those rare instances where outcomes research has been attempted, the methodologic quality has been variable.

This was the background for the workshop titled Methodological Issues in Diagnostic Clinical Trials: Health Services and Outcomes Research in Radiology, sponsored jointly by the U.S. Public Health Service's Office on Women's Health, the National Cancer Institute, and the American College of Radiology's Commission on Research and Technology Assessment, and held March 15, 1998, in Washington, DC. The goals of this workshop were threefold: *(a)* to review state-of-the-art methodologic issues in diagnostic clinical trials, with an emphasis on breast cancer; *(b)* to develop a research agenda focused on the design and modeling of clinical research for analysis of patient outcomes and cost-effectiveness; and (c) to produce a document that will guide program development. This report focuses on the recommendations from each of the workshop's eight sessions.

SESSION 1: CRITICAL REVIEW OF PAST OUTCOMES RESEARCH IN RADIOLOGY

The most commonly used study designs include randomized clinical trials (screening and diagnostic), cohort studies, case-control studies, uncontrolled case studies, and decision analysis studies. Each kind of study has its strengths and weaknesses. All can incorporate health services and outcomes research directly or indirectly. While what constitutes health services and outcomes research in radiology is narrowly defined (eg, levels 4-6 of Pryback and Thornbury's tiered efficacy model [1]), the level of rigor of analyses published in the radiology literature is variable. This variability is a function of the lack of formal training in health services and outcomes research and limited dissemination of fundamental principles. Health services and outcomes research in radiology is also impaired by the complexity of the outcomes component. Specifically, "outcomes" may refer to patient outcomes, where there is a need to separate the effect of diagnosis from treatment, provided there is effective therapy; alternatively, it may relate to intermediate outcomes, measuring the effect on treatment planning that is more proximate to imaging. Other outcomes that warrant consideration include the direct effect of imaging on patient health and health-related quality of life measures. Unfortunately, not all methods for incorporating outcomes into radiology re-

Acad Radiol 1999; 6(suppl 1):S3-S7
' From the U.S. Department of Health and Human Services, U.S. Public Health Service, 200 Independence Ave SW, Hubert H. Humphrey Bldg, Washington, DC 2020I. **Address reprint requests to** S.M.A.

search are mature, nor have they been validated. Another critical issue facing health services and outcomes research is patient selection, with its potential for bias and confounding.

Given the above limitations, several key areas require further development in order to advance and disseminate health services and outcomes research in radiology: methods for diagnostic randomized clinical trials, radiology-specific health status measures, and validation of intermediate outcomes and decision analysis modeling.

SESSION 2: BARRIERS TO CONDUCTING OUTCOMES RESEARCH ON MEDICAL IMAGING IN THE PAST AND HOW THE BARRIERS AN BE COVERCOME

Barriers to conducting outcomes research in medical imaging can be broadly divided into two categories, institutional and external. Institutional barriers are manifold and include a lack of commitment to supporting health services and outcomes research in radiology, absent or cursory training for health services and outcomes research in board-certified radiology residencies and beyond, and, until recently, no academic incentive for conducting this type of research. Additionally, health services and outcomes research has traditionally been performed by medicine and surgery with little crossover or dialogue with radiology departments. On the external side, there are limited resources for health services and outcomes research. In instances where resources may exist, the rationales for pursuing health services and outcomes research are ambiguous, particularly for commercial entities, such as pharmaceutical companies, imaging equipment manufacturers, and managed care organizations. Therefore, the products of these external efforts, including types of results obtained and actions they may produce, reflect competing interests in the marketplace and do not necessarily reflect the societal perspective. This, in turn, can lead to difficulty in generalizing from results. Overriding both institutional and external barriers is the absence of a common vocabulary and a common method of recording data. This lack impedes sharing of information and the validation of results.

Potential solutions for addressing these limitations focus on a commitment to support health services and outcomes research at the residency level, including educational programs and appropriate funding opportunities. This could be mandated by the American Board of Radiology and/or sanctioned by the American College of Radiology. Similarly, academic advancement for health services and outcomes research investigators in radiology will be realized (and thereby validated) only when research of this nature is routinely published in mainstream radiology journals. Additionally, a practical, multidisciplinary approach to health services and outcomes research from the clinical perspective (emphasizing the patient rather than the test) must be adopted. This approach will highlight the trade-offs between the demand for imaging and appropriate utilization and emphasize the importance of posing the appropriate research question. A common vocabulary must be developed if the multidisciplinary approach is to succeed not only at the originating institution but also beyond it.

Addressing external barriers is more problematic. Certainly, increased funding to support health services and outcomes research is needed, but the problem of whose perspective remains (eg, societal vs patient vs health care provider) and is not easily solved. To promote medical informatics, managed care organizations and other industry representatives may be able to share knowledge across different health care providers and manufacturers. Additionally, federal, academic, and industrial partnerships for developing a shared database and lexicon should be encouraged.

SESSION 3: UTILIZATION OF DIAGNOSTIC TESTS – ASSESSING AND APPROPRIATENESS

Practice guidelines are currently being developed in an effort to improve the quality of care and ensure appropriateness. These guidelines may also be used to profile providers and drive payment and patient care decision making. Unfortunately, the "science" behind these guidelines may be limited or nonexistent. Specifically, the types of information used to create appropriateness criteria have not been agreed on; there are no rules governing the inputs, nor have the outcomes been standardized. In many cases, the prerequisite information is lacking. Even where information is available, the quality varies and needs to be characterized.

A research agenda was outlined, with the recognition that appropriateness criteria are and will be used for a variety of purposes. In general, research should define the operating characteristics of guidelines for each specific purpose, the potential domains that might be in common across purposes (guidelines), and the measures of outcomes to be included in the various guidelines. In the interim, methods to improve existing "immature" guidelines must be established and validated, allowing data to be collected and analyzed concurrently. Additionally, a new gen-

eration of statistical models is needed to assess the appropriateness of diagnostic test utilization. These models must be sophisticated enough to account for differences in study design. Similarly, quality control statistical methods require modification to be applicable to health care. To facilitate research synthesis studies, existing databases need to be evaluated. Specifically, work should focus on statistical correlates to confer improved power to smaller, higher-quality databases and improved quality to larger, more "vague" administrative databases.

Last, research is needed to study the effect of socioeconomic and institutional structures on the provision of diagnostic imaging services, in order to determine the effect on current utilization and guideline construction and to provide guidance on infrastructure improvement, so that appropriateness criteria will be most effectively implemented.

SESSION 4: INTERMEDIATE OUTCOMES-DIAGNOSTIC AND THERAPEUTIC EFFECT

Intermediate outcomes are critical for rapidly and efficiently evaluating new technologies and for separating the effects of diagnosis and treatment. The current construct of intermediate outcomes is derived from the efficacy hierarchical model. Specifically, diagnostic thinking efficacy measures (level 3) include assessment of the extent to which the information rendered helps in diagnosis, changes in differential diagnosis probability distribution, and changes in physicians' diagnostic certainty, whereas therapeutic efficacy measures (level 4) capture the effect of imaging information on physicians' treatment decision making (eg, instituting new treatment, continuing current treatment, or avoiding treatment). Patients' utility assessments for intermediate outcomes are addressed by present and future health perceptions, uncertainty about disease prognosis, level of anxiety, quality of life perceptions, and psychological state. These notions may be quantified by using willingness to pay, time trade-off, and standard gamble methods. A variety of questionnaires and other psychometric instruments provide quantitative measures for gauging perceptions about health status, including anxiety, discomfort, and inconvenience.

In considering how to expand on the existing uses of intermediate outcomes, several key points were raised. Specifically, intermediate outcomes should be incorporated earlier in the emergence of a new technology, at or before the diffusion phase. Intermediate outcomes also need to be integrated into current information systems to enhance data collection and access. Ideally, there needs to be a mechanism to make databases accessible via the World Wide Web, provided confidentiality issues are addressed and solved. To ensure, or at the very least test for, generalizability of intermediate outcomes, a multidisciplinary/multiinstitutional approach is necessary, including a spectrum of health care provider environments (eg, tertiary and community hospitals).

SESSION 5: TRADITIONAL HEALTH OUTCOMES IN THE EVALUATION OF DIAGNOSTIC TESTS

Technology assessment typically proceeds in three phases. The first phase characterizes technical performance (eg, spatial resolution). The second phase quantifies diagnostic accuracy (eg, accuracy, sensitivity, specificity, predictive values, and measures of receiver operating characteristic curve height) and is associated with image interpretation. The third phase focuses on the final outcomes of interest, including clinical, economic, and social end points. Traditionally, diagnostic test evaluation proceeds sequentially, beginning with phase 1 and ending with phase 3. However, even with limited early data (phase 1 or 2), innovators of new technology must think of downstream health outcomes. This concept was reinforced by the experts' unanimous endorsement that "radiology cannot choose not to consider health outcomes in the evaluation of diagnostic tests." Clinical trials must address specific issues, but methods for diagnostic trials need to emphasize outcomes. Furthermore, a link should be forged between investigators addressing new methods for evaluation of diagnostic tests and reviewers at regulatory agencies.

SESSION 6: HEALTH PROFILES AND QUALITY OF LIFE

Quality of life measures have been variably incorporated into clinical trials, either as primary end points or in combination with clinical end points. Components of clinical quality of life studies may include explicit a priori hypotheses linking medical care and health-related quality of life (ie, interpretation of results in light of a hypothesized relationship to medical care), selection of sensitive and specific quality of life measures, appropriate time interval for assessment of quality of life changes, avoidance of aggregation (to ensure that disease groups are maintained), and adjustment for baseline differences in case mix.

Quality of life measures have not found widespread use in diagnostic clinical trials. No one measure is the sine qua

non for health status determinations, and a battery of measures may be more appropriate. Specifically, generic measures allow for comparability, whereas disease-specific measures are more responsive to change. Satisfaction measures may also be important. There is increasing interest in assigning quality of life measures to the therapeutic value of a diagnostic test. For example, a negative result and the reassurance it may confer are a legitimate outcome. This is especially true in breast imaging, where a test result, especially a true-negative or a false-positive mammogram, may have a profound effect on an individual. To date, however, these issues have been largely ignored in radiology research, eclipsed by determinations of accuracy, sensitivity, and specificity. Another area where therapeutic value quality of life measures may have merit is the incorporation of "willingness to pay" or patient preference measures into economic analyses (eg, cost utility analyses).

SESSION 7: MEASURING AND INCORPORATING PATIENT PREFERENCES AND UTILITIES IN THE ASSESSMENT OF DIAGNOSTIC TECHNOLOGY

Along with incorporating patient preferences and utilities into diagnostic imaging technology assessment, add-ons to current technology assessments are desirable. For example, preliminary decision modeling can look at the extent to which patient preference drives cost-effectiveness. Similarly, researchers could consider possible short-term preference issues, such as the utility of the diagnostic technology (see the discussion above of false-positive and true-negative mammograms). This approach provides insight into what happens during a clinical trial.

Perhaps the most robust mechanism to supplement current technology assessments is a multipronged approach to ascertaining preferences for material outcomes *beyond* the scope of the trial. Specifically, charting the utility of current health among trial subjects with the health state of interest, as well as utilities for imagined health states of interest, would be important for future decision analysis modeling. Likewise, a preference classification system among patients with the health state of interest could help in determining utilities.

Four research areas were proposed for measuring and incorporating patient preferences and utilities in health services and outcomes research in radiology. The first is the development of utility assessment instruments, both generic and disease/intervention-specific, with the recognition that there is no gold standard. The second area of research emphasizes the basic epidemiology of utility assessment and aims to assess patients' utility of particular diagnostic tests, provide insight into the role of patient preferences in later technology assessment, and provide a catalog of utility inputs for cost-effectiveness analyses. The third area of research is the development of standardized measures of reassurance. These measures need not be specific to diagnostic technologies; however, the question of whether to include them in cost-effectiveness analyses must be frankly discussed. Finally, the fourth research area is the quest for improved methods of assessing utilities for unfamiliar outcomes.

SESSION 8: ESTIMATING COSTS AND COST-EFFECTIVENESS FOR DIAGNOSTIC TECHNOLOGIES

A research agenda for funding institutions, such as the National Institutes of Health, should give priority to cost-effectiveness research based on the likelihood of reducing the future burden of disease. For example, support should be targeted at research that aims to establish the link between diagnostic technologies and human welfare (eg, survival, quality of life, anxiety reduction, cost). Another priority is to design studies that consider all relevant parties (eg, patient, provider, payer, regulatory agencies) and build in clinical and economic end points to address their needs.

There is also a pressing need for a new generation of models. These models should (a) lend credibility to the decision of whether to conduct a randomized clinical trial, *(b)* optimize randomized clinical trial study design, and (c) explore alternative approaches to cost-effectiveness analysis based on secondary data. This new generation of models will reflect improvements in current modeling methods (eg, progression of disease, parameter estimation, uncertainty analysis, multiattribute sensitivity analyses, and model validation).

CONCLUSION

By the end of the conference, six needs essential to the methodologic advancement of diagnostic radiologic clinical trials had emerged: (a) a multidisciplinary approach to health services and outcomes research in such trials; (&) intermediate outcomes to facilitate evaluation and implementation of new imaging technologies; (c) a new generation of models to aid research synthesis studies, attend to preferences, optimize/alter traditional randomized clinical

trial design, and assess appropriateness; *(d)* measures (quality of life and preferences) to assess the therapeutic value of a diagnostic radiologic test; *(e)* improved shared clinical databases and a common lexicon that are easily accessed; and (/) better education in fundamental methods early in post-graduate radiology training programs.

ACKNOWLEDGEMENTS

The authors express their gratitude to the following individuals, whose support made this conference possible: Wanda K. Jones, DrPH, Deputy Assistant Secretary for Health (Women's Health), U.S. Public Health Service, Office on Women's Health; Robert E. Wittes, MD, Deputy Director for Extramural Science and Dirctor, Division of Cancer Treatment and Diagnosis, National Cancer Institute; and Daniel Sullivan, MD, Associate Director Diap nostic Imaging Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute. For editorial input, the authors acknowledge Kevin A. Schulman, MD Director of Clinical Economics Research Unit,-____@__ Georgetown University Medical Center.

REFERENCE

1. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11:88-94.